Self-selection with non-equilibrium beliefs: Predicting behavior in a tournament experiment[a]

Tobias Brünner[b]

Abstract

In this study we use level-k thinking and a recently proposed model of non-equilibrium beliefs in dynamic games (PBNLK) to predict behavior in a tournament with self-selection. We find that the combination of level-k and PBNLK predicts both the population of types in the tournament, as well as the mean and variance of efforts better than Nash equilibrium, a static level-k model and other models of non-equilibrium beliefs. Our results show that non-equilibrium beliefs are an important determinant for the decision to compete in a tournament and the performance in that tournament. Moreover, a useful model of non-equilibrium beliefs should allow players to update their beliefs during the course of the competition.

JEL Classification: C72, C92, D02, M52

Key words: level-k thinking, NLK, Bayesian updating, tournament, experiment,

competition

[a]This work was supported by the German Research Foundation (DFG) under the title BR 4674/1-1. I thank Dan Levin and Luyao Zhang and participants at the FLEX 10-year anniversary conference at Goethe University Frankfurt for valuable comments. I am especially grateful to Marie Claire Villeval for providing the data from the experiments of Eriksson et al. (2009) and for her valuable comments.

[b]Lincoln International Business School, Brayford Pool, LN6 7TS, Lincoln, UK. E-mail: tbruenner@lincoln.ac.uk.

1 Introduction

In modern labor markets, there is considerable variety in the compensation schemes offered to employees both between firms and within firms. These compensation schemes include fixed wages, piece rates, bonuses and relative performance pay such as tournaments.[1] One of the firms' key objectives when implementing a payment scheme is to screen the pool of candidates in the recruitment or promotion process. To achieve this objective and attract applications from the most suitable candidates, firms need a good understanding of how workers choose between payment schemes. As a result, there is a growing literature in labor economics analyzing workers' self-selection into different compensation schemes and the impact on their subsequent performance. Experimental research shows that, holding ability constant, gender, risk aversion, social preferences and the perceived probability of winning affect self-selection into tournaments (Niederle and Vesterlund, 2007; Eriksson et al., 2009, "ETV"; Dohmen and Falk, 2011 and Balafoutas et al., 2012). Moreover, workers that have self-selected into a tournament are more productive than workers that have been assigned to the tournament (ETV and Balafoutas et al., 2012). The findings of these experiments have been confirmed in real labor markets (Flory et al., 2015); and behavior in this type of experiment has predictive power for real labor market entry decisions (Buser et al., 2014).

This study contributes to this literature by focusing on the role of workers' beliefs about their opponents' behavior. In particular, we analyze the usefulness of models of non-equilibrium beliefs for predicting the performance of a tournament with self-selection. ETV's tournament experiment offers ideal conditions for this endeavor.

[1] A survey of 97 UK-based organizations found that 48% of organizations use performance-related rewards schemes. To measure performance 58% of organizations rely on individual performance ratings (like bonuses or piece rates) and 39% use relative performance ratings (like tournaments). Note that these two groups are not mutually exclusive (Bailey et al., 2017).

The experiment consists of two treatments in a between-subject design. The Benchmark treatment is a standard two-player tournament. In the Choice treatment, subjects can self-select into a piece rate compensation scheme or the tournament scheme. In the first step, we estimate the population level frequencies of the level-k model for the Benchmark treatment. In the second step, we use these population level frequencies to predict the outcome of the Choice treatment. Level-k thinking does not explicitly allow players to update their beliefs over the course of a dynamic game. Levin and Zhang (2019) propose a new model of non-equilibrium beliefs (NLK) that bridges Nash equilibrium and level-k thinking. In an extension of their model to dynamic games (PBNLK) players update their beliefs about what type of opponent they face during the course of the game. We find that the combination of level-k and PBNLK accurately predicts (i) the population of types that self-select into the tournament; (ii) the updating of beliefs after the self-selection stage; and (iii) mean and variance of effort in the tournament stage.

The findings of this paper have implications beyond the self-selection decision in labor markets. A good solution concept should help us make sense of the behavior that we observe. But for a solution concept to be useful for policy makers and mechanism designers, it should also have predictive power for future behavior, even when details of the game are modified.[2] Levin and Zhang (2019) have already shown that their PBNLK solution concept explains behavior in centipede games better than Nash equilibrium and level-k thinking. The present study finds that in a tournament context, PBNLK has the greatest out-of-sample predictive power in a modified version of the tournament among all the alternatives considered. Thus, PBNLK and its updating of beliefs during the course of a game seem to be important

[2]Gabaix and Laibson (2008) argue that economic research does not emphasize predictive precision as much as the natural sciences and they "hope that economists will close this gap. Models that make weak predictions (or no predictions) are limited in their ability to advance economic understanding of the world" (p. 296). Studying the performance in out-of-sample predictions also helps to avoid model overfitting.

concepts that promise to be valuable for the design of new policies and deserve more attention in future research.

2 Rank-order tournaments

Rank-order tournaments are introduced by Lazear and Rosen (1981). Two players compete in a tournament. Each player $i \in \{1, 2\}$ chooses an effort level x_i, which results in output $y_i = x_i + \varepsilon_i$. The output shocks ε_i are independently distributed according to a uniform distribution on the interval $[-a, a]$. The player with the highest output receives prize W, the other player receives w, with $W > w$. Players incur costs that are convex and increasing in their own effort choice. Following most experimental studies, we assume a quadratic cost function $C(x_i) = cx_i^2$.

Player i chooses the effort level x_i that maximizes her expected utility,

$$EU_i(x_i, x_{-i}) = G_i(x_i, x_{-i})\,(W - w) + w - cx_i^2\,, \tag{1}$$

where G_i denotes the probability that player i has the highest output.

For $W - w < 8\,c\,a^2$, there is a unique symmetric Nash equilibrium in which both players choose

$$x_1^* = x_2^* = \frac{W - w}{4ac}. \tag{2}$$

Experimental studies of rank-order tournaments, starting with Bull et al. (1987), commonly find that average effort levels are reasonably close to the Nash equilibrium prediction, but the variance in efforts is huge (see Dechenaux et al., 2015, for a survey of this literature.).

ETV's experiment uses the following parameter values: $W = 96, w = 45, a = 40$ and $c = \frac{1}{150}$, which induce an equilibrium effort of 48. The Choice treatment differs

Table 1: Experimental results: Effort levels in tournament

Experiment	Mean	Variance	Equilibrium effort
ETV Benchmark	53.3	652.3	48
	(rounds 1-20)	(rounds 1-20)	
ETV Choice	61.6	258.2	48
	(rounds 1-20)	(rounds 1-20)	

from the Benchmark treatment in that subjects have the choice to opt out of the tournament and work for a piece rate instead. The piece rate is set such that (i) zero effort leads to a fixed payment of 45 (equal to the loser's prize in the tournament) and (ii) the maximum expected utility of a risk neutral player is the same as the expected utility in the Nash equilibrium of the tournament. Therefore, risk neutral players are indifferent between the piece rate and the tournament and those who choose the tournament should again exert the equilibrium effort of 48. In both treatments, subjects played 20 rounds with random rematching after each round.

ETV's results are summarized in Table 1. In the Benchmark treatment, average effort is 53.3, which is slightly above the equilibrium prediction. The more striking feature is the huge variance in efforts of 652.3. In the Choice treatment, the average effort in the tournament is higher (61.6) and the variance in effort is lower (258.2) than in the Benchmark treatment. Balafoutas et al. (2012) find a similar treatment effect for average efforts using a real-effort task and a within-subject design.

The huge variance in efforts, especially in the Benchmark treatment, and the change in the distribution of efforts when moving from the Benchmark treatment to the Choice treatment cannot be reconciled with the Nash equilibrium prediction. In the following, we will explore if models of non-equilibrium beliefs can explain the variance in effort and predict the treatment effect. This is motivated by Balafoutas et al. (2012)'s finding that the expectation of winning the tournament, which is derived from the beliefs about other subjects' effort choices, is the most significant explanatory variable for the choice of the tournament.

3 Level-k and NLK

Level-k thinking relaxes the assumption that players' beliefs about their opponents' strategies coincide with the Nash equilibrium strategies. Instead, beliefs are modeled by a hierarchy of beliefs. The key issue in any application of level-k thinking is the specification of the starting point of this belief hierarchy, the non-strategic level-0 type ($L0$). A common choice for $L0$ is a uniform distribution over all non-dominated actions. In rank-order tournaments all effort levels above the equilibrium effort are dominated. As $L1$ and all higher types will never play dominated strategies, a level-k model that builds on a $L0$ specification that rules out dominated strategies cannot describe the vast amount of effort choices above the equilibrium level.[3]

Thus, a $L0$ specification that puts substantial mass on actions above the equilibrium level seems necessary in rank-order tournaments. Other than that, the qualitative results of this paper are not very sensitive to the choice of $L0$. We assume that $L0$ is uniformly distributed between the equilibrium effort level, 48, and the upper bound of the effort space, 100.

In the first stage of the Choice treatment, in which players choose between the piece rate and the tournament scheme, we assume that $L0$ chooses randomly between the two schemes.

[3]Bernard (2010) makes a related point in the context of using level-k thinking to explain experimental evidence of contests.

3.1 Level-k

Level-k thinking assumes that a player believes that all other players' strategic sophistication is exactly one level below her own level of sophistication. Thus, $L1$ believes that her opponent in the tournament is $L0$. The best response to a uniform distribution on the interval $[48, 100]$ is to choose an effort level of 10.[4] $L2$ believes her opponent is $L1$ and, thus, chooses the best response to an effort level of 10, $L3$ chooses the best response to $L2$'s effort choice and so on. The level-k strategies for ETV's Benchmark treatment are summarized in Table 2.

$L1$ exerts low effort and does not expect to win the tournament. Thus, her expected utility of 46.3 is close to the loser's prize, $w = 45$. Under the piece rate scheme, a player can reach an expected utility equal to the expected utility in the Nash equilibrium, 55.1, independently of the other player's strategy. Therefore, when given the choice between the piece rate and tournament scheme, $L1$ will choose the piece rate. $L2$ (and all types above $L2$) expect to play against an opponent who chooses a lower effort level than their own. Consequently, they have a greater probability of winning the tournament than in Nash equilibrium, while exerting less effort than the equilibrium effort level. Thus, $L2$ and higher types prefer the tournament to the piece rate.

3.2 NLK and PBNLK

Levin and Zhang (2019) propose a generalization of level-k, called NLK, that allows a player to believe that other players are equally sophisticated as she is. In NLK, a NLk player believes that with probability λ she plays against a naive opponent (i.e. a $NL(k-1)$ type) and with probability $1-\lambda$ her opponent is of the same type

[4] The derivation of the level-k strategies is provided in the Appendix.

(i.e. a NLk type). For $\lambda = 1$, NLK is identical to level-k thinking; for $\lambda = 0$, NLK is identical to a Nash equilibrium analysis.[5]

To keep the number of parameters to be estimated small and to avoid ambiguity, we assume that $\lambda \to 1$. Consequently, NLK is identical to level-k in ETV's Benchmark treatment. This assumption is not critical for our main results. What is important is that $NL1$-types choose a low effort level compared to the Nash equilibrium effort and that their expected utility is lower than in Nash equilibrium.[6]

Furthermore, Levin and Zhang (2019) extend their NLK concept to dynamic games. Perfect Bayesian NLK (PBNLK) requires that players update their prior beliefs about other players' sophistication, λ, using Bayes' rule. We use PBNLK to predict behavior in ETV's Choice treatment. In the first stage of the Choice treatment, subjects decide between the piece rate scheme ($d = 0$) and the tournament ($d = 1$). The above analysis shows that $NL1$ prefers the piece rate scheme, i.e., the probability that a player chooses the tournament given that she is $NL1$ is zero, $p(d_{-i} = 1|NL1) = 0$. All higher types prefer the tournament, i.e., $p(d_{-i} = 1|NLk) = 1$, for $k \geq 2$. Thus, when a $NL2$ player learns at the beginning of stage two that she is paired with another player to play the tournament, she will update her prior belief, λ. The posterior belief of playing against an opponent one level below her own level is now given by

$$p^{NL2}(d_{-i} = 1) = \frac{\lambda p(d_{-i} = 1|NL1)}{\lambda p(d_{-i} = 1|NL1) + (1 - \lambda) p(d_{-i} = 1|NL2)} = 0. \qquad (3)$$

Thus, in the second stage of the Choice treatment, $NL2$ types believe they play against an opponent that is as sophisticated as they are, and they will play the Nash equilibrium. Consequently, all higher types will best respond by also choosing

[5]Levin and Zhang (2019) primarily focus on the case where $k = 1$, but they discuss the case with $k > 1$ as an extension.

[6]Applying NLK to a common value auction, Levin and Zhang (2019) find "that for inexperienced bidders (using the first 18 periods), the most accurate prediction of BNLK is with $\lambda = 1$" (p. 26).

Table 2: Predicted effort levels for the different level-k/NLK types. Note: For the NLK model we assume $\lambda \to 1$, such that NLK is equal to level-k in the Benchmark treatment. For ETV's Choice treatment we use PBNLK.

	$L0$	$(N)L1$	$(N)L2$	$(N)L3$	EQ
ETV Benchmark	$[48, 100]$	10	34	43	48
ETV Choice	$[48, 100]$	piece rate	48	48	48

the Nash equilibrium strategy.

Table 2 summarizes the effort choices of the different level-k/NLK types and the Nash equilibrium (EQ) for ETV's experiment. The table reveals that level-k thinking, together with PBNLK, has the potential to explain the general pattern in ETV's experiment. Within a population where all types ($LO, L1, L2, L3$ and EQ) are represented, the variance of efforts will be large in the Benchmark treatment. In the Choice treatment, $NL1$ will not participate in the tournament and $NL2$ and $NL3$ exert higher effort than they would in the Benchmark treatment. Both of these effects lead to higher average effort and lower variance than in the Benchmark treatment.

4 Estimation of types

To categorize the subjects in ETV's experiment into the different level-k, NLK or equilibrium types, we closely follow the mixture model approach pioneered in Stahl and Wilson (1994). In this mixture models approach players make logistic errors. Let k denote the type of a player (e.g. $NL1$, but k can also represent equilibrium beliefs). Player i's observed effort choice in the t-th round is x_{it} and her corresponding expected utility given her belief type k is $S_k(x_{it})$. The probability of

Table 3: Subject-specific type classification

	Benchmark Model 1	Benchmark Model 2	Choice predicted	Choice
Type (k)	π_k	π_k	π_k	π_k
$L0$	30.0%	30.0%	37.5%	26.7%
$(N)L1$	20.0%	20.0%	0%	2.2%
$(N)L2$	13.3%	6.7%	$\sim EQ$	$\sim EQ$
$(N)L3$		30.0%	$\sim EQ$	$\sim EQ$
EQ	36.7%	13.3%	62.5%	71.1%
Log-likelihood	-2581.39	-2578.73		-2251.52

effort choice x_{it}, if subject i is of type k, is then given by

$$Pr(x_{it}|k, \alpha_i) = \frac{exp(\alpha_i\, S_k(x_{it}))}{\int_0^{100} exp(\alpha_i\, S_k(e))de}, \tag{4}$$

where α_i denotes the precision of subject i. As $\alpha_i \to 0$, subject i's effort choices are uniformly distributed on the interval $[0, 100]$; as $\alpha_i \to \infty$, player i always plays exactly the best response.

Let π_{ik} denote the probability that player i is of type k, with $\sum_{k=1}^{K} \pi_{ik} = 1$. For each individual $i = 1, ..., N$ we find the values $(\pi_{i1}, ..., \pi_{iK}, \alpha_i)$ that maximize the likelihood

$$\sum_{k=1}^{K} \pi_{ik} \prod_{t=1}^{T} Pr(x_{it}|k, \alpha_i). \tag{5}$$

Using the estimates of the π_{ik}'s, we can classify subjects into different types and finally obtain the population level frequencies, π_k.

4.1 Results

Table 3 presents the results of the classification for ETV's experiment. Model 1 includes $L0, L1, L2$ together with an equilibrium type (EQ). In addition to all the types in Model 1, Model 2 also includes $L3$. In the Benchmark treatment, Model 1 estimates a high proportion of EQ (40.0%) and $L0$-types (30.0%). But there are also $L1$- and $L2$-types. Model 2 estimates similar proportions of $L0$ and $L1$, but some subjects that have been classified as $L2$ or EQ are now classified as $L3$. A likelihood ratio test, however, shows that adding the $L3$-type does not significantly improve the model.

ETV find convergence of effort choices towards the equilibrium effort level in the Benchmark treatment. As level-k is intended to describe initial play, we redid the classification using only the first ten periods of the experiment. The results are very similar: For 80% of subjects the classification does not change; the subjects that are classified differently do not follow a particular pattern, so that the overall composition is hardly affected.

Another concept that is commonly used to explain behavior in experiments is the quantal response equilibrium (QRE, McKelvey and Palfrey, 1995). In a QRE, players make mistakes, the probability of making a given mistake being decreasing in the cost of that particular mistake. More specifically, in a logistic QRE, the probability of choosing effort x_{it} is given by equation (4). Unlike level-k, however, players are identical and hold consistent beliefs about their opponents' effort choices, i.e., player i believes player j's effort choices follow the distribution implied by equation (4). Dutcher et al. (2015) finds that QRE makes better comparative statics predictions than Nash equilibrium in a tournament experiment. We fit a logistic QRE model to ETV's Benchmark data and find a maximum likelihood estimate of $\alpha = 0.0089$ for the precision parameter and a log-likelihood value of -2767.42. This log-likelihood value is much lower than those for Model 1 and Model 2, but the QRE model has fewer parameters because players are assumed to be symmetric and have the same precision parameter. Using the Bayesian Information Criterion to correct for the difference in estimated parameters, we find that Model 1 (BIC= 5412.8) and Model

2 (BIC= 5490.8) explain the data better than the QRE model (BIC= 5537.6).

4.2 Predicting the Choice treatment

The population level distribution obtained for the Benchmark treatment, together with the belief updating rule of PBNLK, can be used to make an out-of-sample prediction for the Choice treatment which can then be evaluated using the experimental results from the Choice treatment.

The third column of Table 3 shows the predicted distributions of types in the tournament of the Choice treatment. The 20% $(N)L1$-types from Benchmark Models 1 and 2 will choose the piece rate scheme and, thus, there will be no $(N)L1$-types in the tournament of the Choice treatment. $NL2$ and $NL3$ types will update their beliefs about their opponent's sophistication based on equation (3) and behave like equilibrium types.

Column 4 of Table 3 reports the type classification results for the tournaments in the Choice treatment.[7] The estimated population level distribution for the Choice treatment is reasonably close to the predicted distribution in column 3. Most strikingly, only one out of the 45 subjects that chose the tournament at least five times is classified as $L1$.

The out-of-sample prediction for the Choice treatment, based on the proportions of types shown in column 3 of Table 3, implies an average effort of 58.8 and variance of 259.9. These predictions are close to the actual average (61.6) and variance

[7] For the type classification we consider only subjects who have participated in at least five tournaments. This reduces the number of subjects we classify from 60 to 45. The results are not sensitive to this restriction. For the 58 subjects that chose the tournament at least three times, the estimated population level distribution is $L0$: 27.6%, $L1$: 1.7% and EQ : 70.7%.

Table 4: Accuracy of predicted effort levels

	predicted mean	predicted variance	$RMSE$	MAE
PBNLK	58.8	259.9	11.3	8.8
Level-k	52.3	327.7	13.3	11.6
Nash	48	0	21.0	17.3
QRE	48.1	813.3	20.6	17.0
CH ($U[48, 100]$)	74.0	233.9	14.0	12.5
$U[0, 100]$	50.0	856.2	19.6	16.0

(258.2) of effort in the Choice treatment. To put the accuracy of this prediction into perspective, we compare it to alternative models.

First, consider a level-k model without updating. Assume that the subjects in the Choice treatment are drawn from the population level distribution estimated for the Benchmark treatment (Benchmark Model 1 in Table 3), but they do not update their beliefs about their opponents' sophistication. More specifically, although $L1$-types choose the piece rate, $L2$-types believe they will compete with a $L1$-type in the tournament. Table 4 shows the predicted mean and variance of effort for the level-k model without updating. Both mean and variance of effort are further away from those observed in the experiment than the PBNLK prediction.

Table 4 also reports the root mean squared error, $RMSE = \sqrt{\text{mean}((x - \hat{x})^2)}$, and the mean absolute error, $MAE = \text{mean}(|x - \hat{x}|)$, for the PBNLK, level-$k$ and Nash equilibrium prediction. (Here, x and $\hat{x}$ are vectors of observed and predicted efforts, respectively, both arranged in ascending order.) Both measures show that PBNLK predicts effort better than the level-k model without updating, which in turn predicts better than Nash equilibrium. We use the Diebold-Mariano test to assess the significance of differences in predictive accuracy (Diebold and Mariano, 1995).[8] All the differences in the predictive accuracy between the PBNLK, level-k and Nash model are significant.

Second, we consider the prediction based on the QRE model discussed in the preceding section. In the QRE model the decision to participate in the tournament does not reveal any information about a player's type or their intended behavior. There is, therefore, no reason to believe that behavior in the tournament during the Choice treatment will be any different from that during the Benchmark treatment. Thus, we use the QRE model with the precision parameter $\alpha = 0.0089$ to predict efforts. Table 4 shows that the prediction of the QRE model is poor; it narrowly beats the Nash equilibrium prediction, although the difference is insignificant, and is significantly behind the PBNLK and static level-k predictions.

Third, we consider the predictions of a Cognitive Hierarchy model (CH, Camerer et al., 2004). Similar to level-k thinking, the CH model consists of a hierarchy of types. For $L0$-types and $L1$-types, the CH model is identical to level-k thinking. $L2$-types, however, believe their opponent is drawn from a population consisting of $L0$ and $L1$-types. Similarly, $L3$-types believe they face an opponent drawn from a population consisting of $L0$, $L1$ and $L2$-types. As in the level-k model, $L1$-types in the CH model will choose the piece rate in the Choice treatment. A $L2$-type will then expect to face a $L0$ opponent if she chooses the tournament. Consequently, the $L2$-type prefers the piece rate. The same is true for the $L3$-type. As a result, in the CH model only $L0$-types choose the tournament in the Choice treatment and, thus, effort choices should be uniformly distributed between 48 and 100. Table 4

[8] The Diebold-Mariano test is usually applied to time series forecasts; we apply it to the current setting by regressing the difference in squared (absolute) prediction errors between two predictions on an intercept with random effects on the subject level. For example, to test if the MAE of the PBNLK prediction is significantly different from the MAE of the Nash prediction, we run the regression $|\hat{x}^{\text{PBNLK}} - x_{it}| - |\hat{x}^{\text{Nash}} - x_{it}| = \beta_0 + v_{it}$, where the composite errors v_{it} include random effects on the subject level and idiosyncratic errors.

shows that this prediction is better than the QRE prediction but worse than the predictions from the PBNLK and static level-k models. All the differences between the accuracy of the CH, QRE, PBNLK and static level-k predictions reported in Table 4 are significant, except the difference between CH and static level-k when using squared deviation, which is only marginally significant (p-value=0.0561).

The final row of Table 4 reports the prediction of random play over the entire action space from 0 to 100. Again, this prediction is significantly less accurate than the predictions from the PBNLK and static level-k models.

5 Conclusions

We find that the combination of level-k and PBNLK accurately predicts (i) the population of types that self-select into the tournament; (ii) the updating of beliefs after the self-selection stage; and (iii) mean and variance of effort in the tournament stage. Of course, our analysis merely shows that the combination of level-k and PBNLK predicts behavior in ETV's experiment better than Nash equilibrium and other models of non-equilibrium beliefs; it does not imply that these are the "right" models. There might be other behavioral drivers that affect behavior in tournaments, in addition to the models of non-equilibrium beliefs considered here. E.g., ETV and Balafoutas et al. (2012) show that risk aversion partly explains the choice between tournament and piece rate and the subsequent effort provision. Note that adding modest levels of risk aversion does not change the qualitative results of our analysis. For $L1$-types low effort means low risk, so adding risk aversion makes reducing effort even more attractive. A risk averse $L2$-type will exert more effort than a risk neutral $L2$-type to increase their chance of succeeding and thereby reducing risk. The same is true for all types above $L2$. Thus, with risk aversion $L1$ will still prefer the piece rate and the Bayesian updating rule in the Choice treatment will be the same as under risk neutrality. All that changes is that $L2, L3$ and equilibrium types will exert higher effort than under risk neutrality.

The accuracy of the predictions of level-k thinking and PBNLK, both qualitatively and quantitatively, provide a strong indication that non-equilibrium beliefs are an important determinant for the decision to compete in a tournament and the subsequent performance in that tournament. Moreover, our results highlight that a useful model of non-equilibrium beliefs should allow players to update their beliefs during the course of the game.

It is worth bearing in mind that the Lazear/Rosen (1981) tournament model is not restricted to labor markets and is often used as an analogy for competitive environments more generally. Acknowledging that non-equilibrium beliefs affect entry into and behavior in competitive environments is valuable. Self-selection on the basis of individual characteristics like productivity, risk aversion or willingness-to-pay are not new in economics. But unlike personal characteristics, beliefs can be changed, often at negligible cost. For example, to increase uptake of a newly introduced competitive scheme, a one-hour training session, attended by all potential participants, that describes the strategic considerations involved in that particular form of competition can bring all participants on (roughly) the same page and thereby reduce the proportion of $L0$- and $L1$-types. Alternatively, providing selected examples of behavior and outcomes from past instances of that type of competition can help to focus beliefs around the equilibrium action. Even if it is not possible to influence people's beliefs, understanding that beliefs matter, helps to have more realistic expectations about the performance of the competitive scheme.

Appendix: Derivation of level-k strategies

$L1$'s effort level

From the perspective of a $L1$-type who chooses effort x_i, the probability of winning the tournament is $\text{Prob}(x_i \geq x^{L0} + \varepsilon_{-i} - \varepsilon_i)$, where x^{L0} follows a uniform distribution between 48 and 100. Let $z \equiv \varepsilon_{-i} - \varepsilon_i$ and $y \equiv x^{L0} + z$. The probability density functions of z and y are then given by

$$g(z) = \begin{cases} \frac{z+80}{80^2} & , -80 \leq z < 0 \\ \frac{80-z}{80^2} & , 0 \leq z \leq 80 \end{cases} \tag{6}$$

and

$$f(y) = \int_{48}^{100} g(y - x^{L0}) \frac{1}{52} dx^{L0},$$

respectively. Thus, $\text{Prob}(x_i \geq x^{L0} + \varepsilon_{-i} - \varepsilon_i) = \text{Prob}(x_i \geq y) = F(x_i)$, where

$$F(y) = \begin{cases} \frac{1}{6}\frac{1}{52}\frac{(y+32)^3}{80^2} & , 0 \leq y < 20 \\ \frac{1}{6}(\frac{52}{80})^2 + \frac{1}{2}\frac{(y-20)(y+32)}{80^2} & , 20 \leq y < 48 \\ \frac{1}{6}(\frac{52}{80})^2 + \frac{7}{40} + \frac{1}{2}\frac{(y-48)(y+60)}{80^2} - \frac{1}{3}\frac{1}{52}\frac{(y-48)^3}{80^2} & , 48 \leq y \leq 100. \end{cases} \tag{7}$$

The expected utility of a $L1$-type is then given by

$$EU_i^{L1}(x_i) = F(x_i)(96 - 45) + 45 - \frac{1}{150}x_i^2.$$

The graph of this expected utility function in Figure 1 reveals that there is a unique interior maximum in the region $0 \leq x_i \leq 20$. By solving the first-order condition $\frac{1}{2}\frac{1}{52}\frac{(x_i+32)^2}{80^2}51 = \frac{1}{75}x_i$, we find that this maximum is at $x_i^{L1} = 10.265$. As ETV's experiment only allows integer effort levels, we round this to 10.

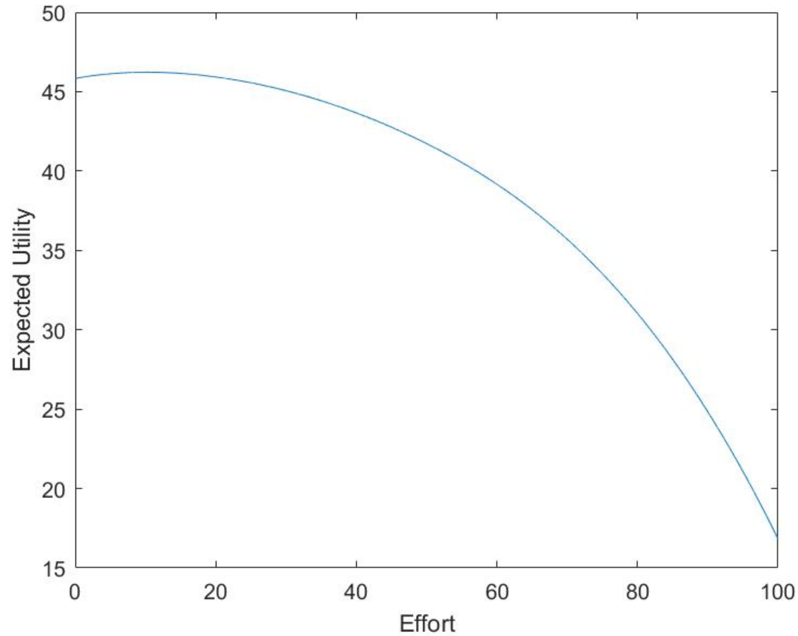

Figure 1: Expected Utility of $L1$-type

Lk's effort level, for $k \geq 2$

To find the level-k effort levels for $L2$ and $L3$, we first derive player i's best response to player j's effort level. If the two players choose effort levels x_i and x_j, the probability that player i has higher output is

$$G_i(x_i - x_j) = \begin{cases} 0 & , x_i - x_j < -80 \\ \frac{1}{2}(\frac{80+x_i-x_j}{80})^2 & , x_i - x_j \in [-80, 0] \\ 1 - \frac{1}{2}(\frac{80-(x_i-x_j)}{80})^2 & , x_i - x_j \in [0, 80] \\ 1 & , x_i - x_j > 80\,, \end{cases} \tag{8}$$

and player i's expected utility is given by

$$EU_i(x_i, x_j) = G_i(x_i - x_j)\,(96 - 45) + 45 - \frac{1}{150}x_i^2. \tag{9}$$

The best response of player i to player j's effort x_j , is then given by

$$x_i(x_j) = \begin{cases} \frac{153}{409}\,(80 + x_j) & , x_j \in [0, 48] \\ \frac{153}{103}\,(80 - x_j) & , x_j \in [48, 80] \\ 0 & , x_j \in [80, 100]\,. \end{cases} \tag{10}$$

Using this best response function, we find that $L2$'s best response to $x_j^{L1} = 10$ is $x_i^{L2} = 34$, and $L3$'s best response to $x_j^{L2} = 34$ is $x_i^{L3} = 43$.

References

Bailey, Marie, Liz Marriott, and Stephen J. Perkins, "Reward management: focus on pay," Survey report, CIPD December 2017.

Balafoutas, Loukas, Rudolf Kerschbamer, and Matthias Sutter, "Distributional preferences and competitive behavior," *Journal of Economic Behavior and Organization*, 2012, *83*, 125–135.

Bernard, Mark, "Level-k reasoning in contests," *Economics Letters*, 2010, *108* (2), 149–152.

Bull, Clive, Andrew Schotter, and Keith Weigelt, "Tournaments and piece rates: an experimental study," *Journal of Political Economy*, 1987, *95* (1), 1–33.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek, "Gender, competitiveness, and career choices," *Quarterly Journal of Economics*, 2014, *129* (3), 1409–1447.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong, "A cognitive hierarchy model of games," *Quarterly Journal of Economics*, 2004, *119* (3), 861–898.

Dechenaux, Emmanuel, Dan Kovenock, and Roman M. Sheremeta, "A survey of experimental research on contests, all-pay auctions and tournaments," *Experimental Economics*, 2015, *18* (4), 609–669.

Diebold, Francis X. and Roberto S. Mariano, "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 1995, *13*, 253–263.

Dohmen, Thomas and Armin Falk, "Performance pay and multidimensional sorting: productivity, preferences, and gender," *American Economic Review*, 2011, *101*, 556–590.

Dutcher, E. Glenn, Loukas Balafoutas, Florian Lindner, Dmirty Ryvkin, and Matthias Sutter, "Strive to be first or avoid being last: An experiment on relative performance incentives," *Games and Economic Behavior*, 2015, *94*, 39–56.

Eriksson, Tor, Sabrina Teyssier, and Marie-Claire Villeval, "Self-selection and the efficiency of tournaments," *Economic Inquiry*, 2009, *47* (3), 530–548.

Flory, Jeffrey A., Andreas Leibbrandt, and John A. List, "Do competitive workplaces deter female workers? A large-scale natural field experiment on job entry decisions," *Review of Economic Studies*, 2015, *82* (1), 122–155.

Gabaix, Xavier and David Laibson, "The seven properties of good models," in Andrew Caplin and Andrew Schotter, eds., *The Methodologies of Modern Economics: Foundations of Positive and Normative Economics*, Oxford: Oxford University Press, 2008, chapter 12, pp. 292–299.

Lazear, Edward P. and Sherwin Rosen, "Rank-order tournaments as optimum labor contracts," *Journal of Political Economy*, 1981, *89* (5), 841–864.

Levin, Dan and Luyao Zhang, "Bridging level-k to Nash equilibrium," 2019. Working Paper.

McKelvey, Richard D. and Thomas R. Palfrey, "Quantal response equilibria for normal form games," *Games and Economic Behavior*, 1995, *10*, 6–38.

Niederle, Muriel and Lise Vesterlund, "Do women shy away from competition? Do men compete too much?," *Quarterly Journal of Economics*, 2007, *122*, 1067–1101.

Stahl, Dale O. and Paul W. Wilson, "Experimental evidence on players' models of other players," *Journal of Economic Behavior and Organization*, 1994, *25*, 309–327.